

Comments on: Modular Regression - A Lego System for Building Structured Additive Distributional Regression Models with Tensor Product Interactions

Goicoa, T^{1,2}

Received: date / Accepted: date

The authors address the important topic of building very general models with interaction terms facing the relevant issue of identifiability. I think the topic is pertinent and timely as models are more and more complicated to describe complex phenomena involving different kind of information. In my opinion identifiability issues should be studied in a clear way so that users can understand what is going on. I congratulate the authors for this stimulating paper.

Constraints vs. reparameterization

The main contribution of the paper is the constrained sampling proposal for identification that conserves matrix sparsity, unlike other proposals such as model reparameterization that increase computing burden. The authors propose a general framework based on orthogonality in function spaces that extend the class of functions that can be conveniently removed, and they clearly establish some of the advantages and disadvantages of both procedures. For example, a reparameterization based on the eigen-decomposition of the precision matrix reduces the dimension of the parameter space whereas constrained sampling does not. On the other hand, reparameterization does not conserve sparsity and constrained sampling does. The distinction between reparameterization and constraints is important as there are situations in which we have

Tomás Goicoa

¹ Department of Statistics, Computer Science and Mathematics. Public University of Navarre.

² Institute for Advanced Materials, INAMAT. Public University of Navarre.

Present address: Campus de Arrosadia, 31006, Pamplona, Spain.

E-mail: tomas.goicoa@unavarra.es

Tel.: +34-948-169199

Fax: +34-948-169204

This is a post-peer-review, pre-copyedit version of an article published in Test.

The final authenticated version is available online at: <https://doi.org/10.1007/s11749-019-00633-x>

relevant side effects and reparameterization may be more convenient. One of such situations occurs in spatial and spatio-temporal disease mapping when the Leroux et al. (1999) CAR prior (LCAR) is considered for the spatial random effect. The LCAR prior is proper (unless the spatial weight $\lambda = 1$) but still there is a confounding issue with the model intercept. Goicoa et al. (2018) show that in this case there is a difference between reparameterizing the model or using constraints. If the model is fitted subject to constraints the variance of the intercept is inflated. In a non-Bayesian analysis the LCAR intercept is always estimated as zero. Retaining the LCAR intercept leads to a variance inflation of the model intercept because a column of ones is an eigenvector of the precision matrix but its associated eigenvalue is not null, unlike the intrinsic CAR prior where the eigenvalue associated to the column of ones is null and the variance inflation does not occur. This phenomenon also happens with the prior proposed by Dean et al. (2001) and later modified by Riebler et al. (2016). Given that using constraints seems to be more convenient in terms of computing burden than reparameterizing, it is relevant to determine in which situations fitting a model with constraints may have undesirable side effects.

Orthogonality in function spaces

Regarding tensor product interactions, orthogonality in function spaces offers the possibility of removing a much wider class of effects from the interaction than using a reparameterization based on the eigen-decomposition of the precision matrix. In general, the effects in the null space of the precision matrix are constant effects, linear effects, and so on, which are easy to interpret. In my opinion it is not clear when or in which settings removing effects beyond those of the null space of the precision matrix will be of interest. For example, the authors comment functional random effects as a special case of the method proposed in the paper, but functional random effects can be also identified via reparameterization or constraints based on the eigen-decomposition of the precision matrix. For example, Goicoa et al. (2017) consider different disease mapping models with area-specific temporal trends, area-specific age curves or area-specific age-time surfaces. Wood et al. (2013) also provide a flexible and intuitive method to construct tensor product smooths based on reparameterization.

Tensor product and type IV interactions

There are two interesting questions related to tensor product and Type IV interactions. I find appealing the way to consider anisotropy in the tensor product interaction through the weight ω and the interpretation as the importance of one prior relative to the other. The authors do not comment if this idea can be extended to the case of, let us say, three dimensional penalties $\mathbf{K} = \frac{1}{\tau_1^2} \mathbf{K}_1 \otimes \mathbf{I}_2 \otimes \mathbf{I}_3 + \frac{1}{\tau_2^2} \mathbf{I}_1 \otimes \mathbf{K}_2 \otimes \mathbf{I}_3 + \frac{1}{\tau_3^2} \mathbf{I}_1 \otimes \mathbf{I}_2 \otimes \mathbf{K}_3$. Including two different

precision parameters in the two dimensional penalty has consequences over the MCMC implementation slowing down the computations. However, other alternatives to MCMC for Bayesian inference can be chosen. Implementing tensor product interactions with different smoothing parameters is possible in INLA (Rue et al., 2009) using the latent model `generic3` in relatively short time (see for example Ugarte et al., 2017).

Finally the authors seem to prefer tensor product rather than Type IV interactions given that the null space of the former are smaller. I agree with that, but if the Type IV interaction is conveniently constrained, differences are small with only a smoothing parameter. The authors mainly focus on aspects related to the null spaces of the precision matrices, but I miss some kind of interpretation about both penalties. Type IV interactions are widely used in disease mapping and they include the relationship between second order neighbours (Knorr-Held, 2000). This implies that area i in time t is influenced by neighbouring areas j in time t , area i in time $t - 1$ and $t + 1$, and finally neighbouring areas j in time $t - 1$ and $t + 1$ (if a RW1 is considered for time), something that seems sensible. Intuitively, this Type IV interaction gives rise to area-specific temporal trends that tend to be similar for neighbouring regions. In tensor product interactions temporal trends in neighbouring areas also tend to be similar, but they do not consider second order neighbours. Is the smoothing very different? Does any of them lead to a greater amount of smoothing than the other?

Acknowledgements

This work has been supported by project MTM2017-82553-R (AEI, FEDER/UE). I would like to thank the Editor for the invitation to discuss this stimulating paper.

References

1. Dean CD, Ugarte MD, and Militino AF. (2001). Detecting interaction between random regions and fixed age effects in disease mapping. *Biometrics*, **57**: 197-202.
2. Goicoa T, Adin A, Etxeberria J, Militino AF, and Ugarte MD. (2017). Flexible Bayesian P-splines for smoothing age-specific spatio-temporal mortality patterns. *Statistical Methods in Medical Research*. DOI: <https://doi.org/10.1177/0962280217726802>.
3. Goicoa T, Adin A, Ugarte MD, and Hodges JS. (2018). In spatio-temporal disease mapping models, identifiability constraints affect PQL and INLA results. *Stochastic Environmental Research and Risk Assessment*, **32**: 749-770.
4. Knorr-Held L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine* **19**: 2555-2567.
5. Leroux BG, Lei X, and Breslow N. (1999). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment and Clinical Trials* (eds Halloran M and Berry D), 179-191. New York: Springer-Verlag.
6. Riebler A, Sørbye SH, Simpson D, and Rue H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, **25**: 1145-1165.

-
7. Rue H, Martino S, and Chopin N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **71**: 319-392.
 8. Ugarte MD, Adin A, and Goicoa T. (2017). One-dimensional, two-dimensional, and three dimensional B-splines to specify space-time interactions in Bayesian disease mapping: Model fitting and model identifiability. *Spatial Statistics*, **22**, Part 2: 451-468.